INDIVIDUALIZATION FEATURE OF HEAD-RELATED TRANSFER FUNCTIONS BASED ON SUBJECTIVE EVALUATION

Satoshi Yairi, Yukio Iwaya and Yôiti Suzuki

Research Institute of Electrical Communication /
Graduate School of Information Sciences, Tohoku University,
2–1–1 Katahira, Aoba-ku, Sendai 980–8577, Japan
{yairi, iwaya, yoh}.riec.tohoku.ac.jp

ABSTRACT

To realize a three-dimensional virtual sound image with a Virtual Auditory Display (VAD), it is important to individualize Head-Related Transfer Functions (HRTFs) for listeners. The purpose of the present study was to establish a fitting method for HRTFs based on a listening test. To this end, a number of sets of virtual images were synthesized using HRTFs of different individuals. The listeners were then asked to choose appropriate virtual sound images located in the intended orbits. To achieve our fitting method of HRTFs by such a subjective evaluation, it is desirable that the same HRTFs can be chosen with stability. In this study, the process used to select a set of HRTFs subjectively was examined in detail, and the features of the individualization of HRTFs by subjective evaluation were investigated. First of all, the process to choose the best of 32 sets of HRTFs by a Swiss-style tournament was repeated ten times, and the regularity of wins in the tournament was examined. As a result, it was understood that the same set of HRTFs is not always chosen and that the individualization method has probability features. The strength of the sets of HRTFs which won the tournament several times was then evaluated. A round-robin comparison with the 130 sets of HRTFs in our HRTF-corpus was repeated twenty times. It was shown that a subjective evaluation itself was also a probability feature. Moreover, the percentage of winning for the set of HRTFs which won the tournament was estimated to be about 15% from the results of the round-robin comparison.

1. INTRODUCTION

Head-Related Transfer Functions (HRTFs) [1] are transfer functions that are related to sound propagation paths from a sound source to a listener's ears. Commonly, Virtual Auditory Display (VAD) techniques are based on convolving HRTFs to a sound source. When HRTFs in a VAD are not fitted to a specific listener, the accuracy of localization is often low and localization errors are produced, typically as front-back confusion [2] [3]. However, measurement of a listener's own HRTFs in all directions requires a special measuring apparatus, a long measurement time, and physical load on a subject. Therefore, it is necessary to estimate the most individualized HRTFs without acoustical measurement.

If the relationship between features of the spectrum shape of HRTFs and the listener's anthropometry is revealed, it is possible to create HRTFs by scanning the listener's body. Though considerable research from such an objective viewpoint has been done [4] [5] [6] [7], there is no established technique.

On the other hand, research has been done to construct an individualization technique from the viewpoint of subjective evaluation. Seeber *et al*. [8] have proposed a subjective selection method consisting of two steps, and Iwaya [9] has proposed a new fitting method called the Determination method of OptimuM Impulse-response by Sound Orientation (DOMISO). In the latter method, listeners are asked to choose the most suitable set of HRTFs from among 130 sets of HRTF data in the corpus [10] based on tournament-style listening tests. He reported that there were no significant differences of the front-back error rate between selected HRTFs and the listener's own HRTFs. The present authors have also investigated the effect of two kinds of virtual orbits used in DOMISO, *i.e*., a horizontal orbit and an upper horizontal orbit (30 degrees above), to control a sound image in the sagittal plane [11]. Sound localization performance was compared between sets of HRTFs selected from those two orbits. Results showed that HRTFs based on the DOMISO selection with the upper horizontal orbit showed a better performance than that with the horizontal orbit. However, the features of subjective evaluation of HRTFs were not discussed in these previous studies. Therefore, in the present study, the subjective evaluation process used in the selection of a set of HRTFs from among many sets was investigated in detail. Points of interest in this study are as follows: 1) Stability of tournament results, 2) Strength of a set of HRTFs selected in a tournament, and 3) Percentage of winning of a selected set of HRTFs in a tournament. Experiments were performed using horizontal orbits to evaluate these points.

2. EXPERIMENT 1: VERIFICATION OF STABILITY OF TOURNAMENTS

2.1. Method

In this section, the stability of DOMISO tournaments is discussed. The sets of HRTFs used in the experiment were selected from our original corpus of HRTFs [10], which includes 130 sets of HRTFs measured with a spherical speaker array installed in an anechoic room at the authors' institute (Fig. 1). HRTFs for sound sources located 1.5 m from the center of the spherical array were measured with an equal interval angle of 5 degrees in the horizontal plane and one of 10 degrees in the median plane (-80, -70, ..., 0, ..., 70, 80:17 of elevation angles). The length of the impulse response of each HRTF was 512, with a sampling frequency of 48 kHz.

The procedure based on DOMISO is as follows:

1. The cepstrum distance between a set of a listener's own HRTFs and each set of HRTFs is calculated, and all 130 sets in our HRTF corpus [10] are sorted based on the distances. In an actual situation in DOMISO, a listener's own

set of HRTFs is not obtained; however, as this experiment aimed at verification of the stability of DOMISO, the listener's own set is used as a reference.

2. Thirty-two sets of HRTFs are chosen, one from every fourth set of 130 sorted sets.
3. The orbit of a sound source which makes a continuous rotation around the listener is prepared using each of the 32 selected sets of HRTFs. As a result, 32 kinds of orbits are generated. In this study, the orbit was a horizontal plane.
4. The outline of this orbit drawn as a virtual sound image is shown to listeners as an illustration before the listening test.
5. Tournament matches are scheduled for these 32 orbits rendered by the 32 sets of HRTFs.
6. In a session, the listener selects one of two orbits that better resembles the illustrated orbit from the point of view of overall features of the sound image. The judgement is based on the listener's total impression formed freely without any concrete instructions. The selected orbit is the winner of the match. The winner then proceeds to subsequent matches.
7. Finally, one set of HRTFs wins the tournament; it is selected as the individualized HRTFs based on subjective evaluation.

A Swiss-style tournament is used with slight modification in DOMISO. In a Swiss-style tournament, a match is scheduled between HRTF sets with the same number of wins. This style avoids the defeat of two strong sets of HRTFs at early stages of the tournament. In the present study, in addition to the original procedure, any set of HRTFs that had lost three times was removed from the tournament. A schematic of the experiment based on the DOMISO procedure is shown in Fig. 2.

Listeners were three young males and two young females with normal hearing. Among the five listeners, thirty-two sets of HRTFs used in the tournament were different. In this experiment, ten different tournaments were performed for each listener. The same 32 sets of HRTFs were used in the ten tournaments, but the initial tournament schedule was different among the ten tournaments. Other experimental conditions are shown in Table 1. The orbit in a horizontal plane shown to the listeners is illustrated in Fig. 3.

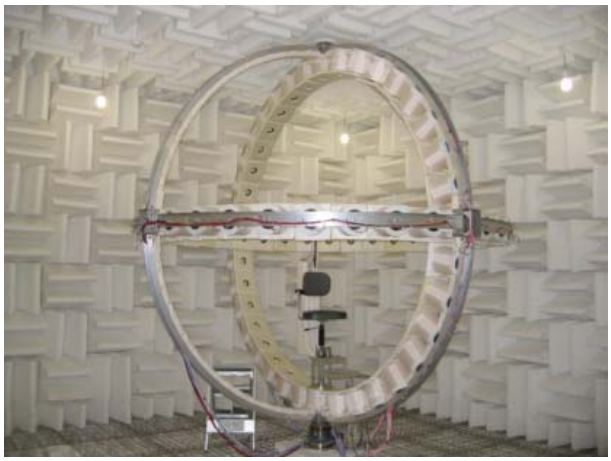


Figure 1: *Spherical speaker array to measure HRTFs.*

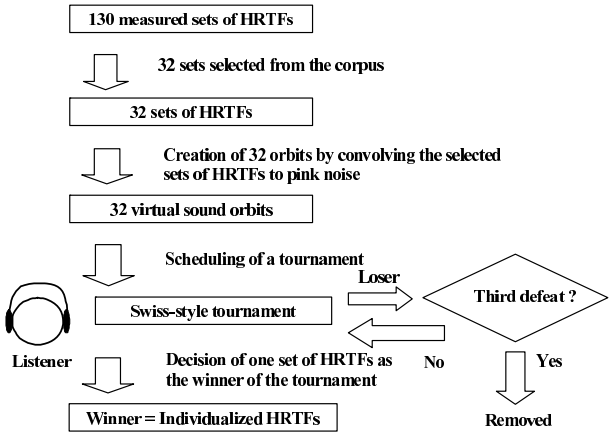


Figure 2: *Schematic of Experiment 1 based on DOMISO procedure.*

Table 1: *Experimental conditions*

Signal	Pink noise
Orbit	Horizontal plane
Method of presenting	Headphones
Listeners	Five adults with normal hearing
Head movement	Restricted

2.2. Results and discussion

Table 2 shows the number of sets of HRTFs which won the tournament from one to three times in ten tournaments for each listener. Strong sets of HRTFs which won the tournament once or more existed for all listeners. However, there was no definitive winner. This indicates that there is more than one adequate set of HRTFs and that several sets performed well in the tested orbit.

Table 2: *Number of HRTFs sets which won the tournament one to three times for each listener*

Listener	Number of wins		
	Three	Two	One
1	1	1	5
2	1	1	5
3	-	2	6
4	1	1	5
5	-	1	8

2.2.1. Details of tournaments

In the experimental results, there were no definitive winners of the ten tournaments. Therefore, we investigated the details of the HRTFs. Table 3 shows the HRTFs which placed from first to eighth in each tournament for Listener 1. The descriptions such as H01 and H02 shown in the table and common to all listeners are

Table 3: *HRTFs with the eight highest ranks in ten tournaments (Listener 1)*

Rank	Trial									
	1	2	3	4	5	6	7	8	9	10
1st	H99*○	H123	H86	H41*	H99*○	H99*○	H44	H5*	H5*	H101
2nd	H123	H86	H7	H5*	H98	H5*	H94*	H31	H41*	H52
3rd	H69	H94*	H5*	H7	H54*	H123	H72	H100	H54*	H99*○
4th	H54*	H41*	H94*	H115	H31	H101	H41*	H99*○	H94*	H100
5th	H102	H39	H99*○	H106	H41*	H54*	H98	H94*	H99*○	H5*
6th	H110	H31	H120	H52	H115	H39	H7	H91	H52	H41 *
7th	H41*	H7	H52	H99*○	H44	H110	H91	H54*	H28	H44
8th	H39	H72	H24	H54*	H5*	H44	H39	H101	H49	H94*

*: sets which placed eighth or higher more than four times in ten tournaments
○: best set from the point of view of averaged rank

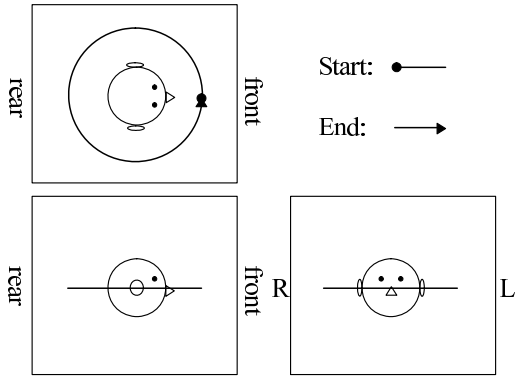


Figure 3: *Illustration of horizontal orbit shown to listeners*.

the labels for sets of HRTFs. Moreover, "*" indicates that the set placed eighth or higher more than four times in ten tournaments. The symbol "○" with H99 indicates that H99 was the best set from the point of view of averaged rank. It should be understood that even though H99 was the best set, it did not place eighth or higher more than twice. Although H99 was the strongest, it did not always win a tournament. It can be concluded that strength is not definitive.

2.3. Consideration employing loss-margin

In the experiment, any set of HRTFs that lost three times was removed from the tournament schedule. However, when the set defeated an opponent with the same strength, the true strength could not be estimated based only on the tournament ranking. Therefore, the strength of three defeated opponents was evaluated. For this purpose, the number of wins of defeated opponents was compared with that of the high-ranked sets in the tournament. The loss-margin S is defined as a ratio between them as follows:

$$S = \frac{w_{lost}}{w_{high}}, \tag{1}$$

where the meaning of w_{lost} and w_{high} are respectively as follows:

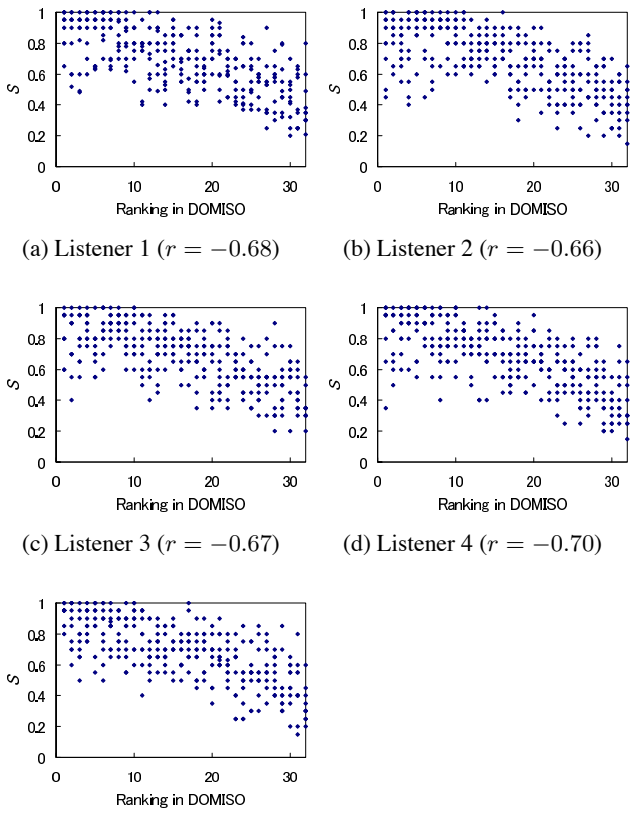
- w_{lost}: the averaged number of wins of the three defeated opponents.
- w_{high}: the averaged number of wins of the three sets of HRTFs which were ranked third place or higher in a specific tournament.

Figure 4 shows the relationship between DOMISO ranking and the loss-margin S of all ten tournaments for each listener. When S is nearly equal to one, it might be considered that the set was defeated by high-ranked opponents in the tournament. Correlation coefficient r is smaller than -0.65 ($p < .01$) for all listeners. However, some sets of HRTFs of lower rank had an S value larger than 0.8. These sets of HRTFs may have been strong, even though they were defeated by an opponent with the same strength.

Let us now review the results of the ten tournaments and the strength of HRTFs using loss-margin S. One or two sets which easily won ten tournaments were chosen and the number of times when the loss-margin of the set was greater than the averaged loss-margin of all tournament participants (k) was calculated. Table 4 shows the results for each listener. The summation of victories (w) and k is also shown in Table 4 as N. The k of all HRTFs excluding H133 is more than or equal to half of (10-w), the number of tournaments which the HRTFs could not win. Therefore, the performance of selected sets of HRTFs was better than that of averaged sets. Furthermore, it can be seen that N is larger than four and most Ns are larger than six. This means that all listeners had sets of HRTFs which were suitable for sound localization.

Table 4: *Strength of HRTFs which won the tournament more than two times*

Listener	HRTFs	#w	#k	N (#w+#k)
1	H99	3	5	8
	H5	2	5	7
2	H133	2	3	5
	H25	2	5	7
3	H26	2	5	7
	H15	2	5	7
4	H93	3	4	7
	H111	2	7	9
5	H29	2	4	6



(a) Listener 1 ($r = -0.68$) (b) Listener 2 ($r = -0.66$)

(c) Listener 3 ($r = -0.67$) (d) Listener 4 ($r = -0.70$)

(e) Listener 5 ($r = -0.70$)

Figure 4: *Relationship between ranking in DOMISO and loss-margin S.*

3. EXPERIMENT 2: VERIFICATION OF STRENGTH BY ROUND-ROBIN

3.1. Method

The previous experiment suggested that each selection based on subjective evaluation is not definitive. The strength of all sets of HRTFs are estimated accurately if round-robin comparison among all sets can be performed many times. However, as it takes about sixty hours to conduct one round-robin comparison, this is not practical. Thus, a round-robin comparison among selected sets and other sets was performed. The 32 sets of HRTFs were sorted according to the number of total wins in Experiment 1 for each listener, and three sets which placed third or higher were chosen as the "best" HRTFs. For comparison, a set of the lowest ranking HRTFs ("worst") and a set of a listener's own measured HRTFs ("own") were used. A total of five sets of HRTFs were subjectively evaluated against all other sets of HRTFs in the corpus. The evaluation was repeated twenty times.

3.2. Results and discussion

The number of wins in these twenty matches was classified into four categories, *i.e.*, 0-5 wins, 6-10 wins, 11-15 wins, and 16-20 wins. The results are shown in Figs. 5-7. These figures respectively indicate "best," "worst," and "own."

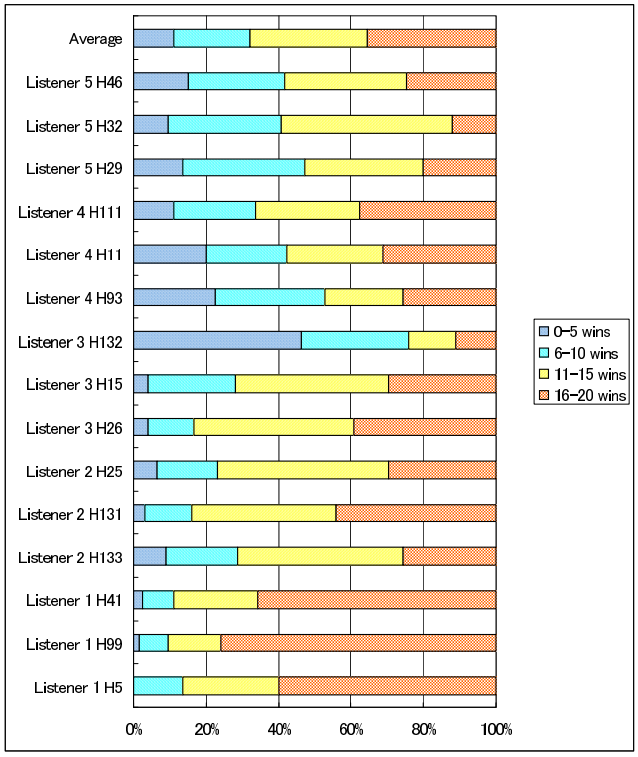


Figure 5: *Classification of all evaluation results in round-robin conducted twenty times ("best" HRTFs).*

In Fig. 5, for sets which won more than 10 out of 20 matches, the percentage for the " best" set was more than 50%, excluding H93 and H132 of Listener 4. In particular, H5, H99, and H41 of Listener 1 won more than 15 times against half the sets of the corpus, while the "worst" set of HRTFs was defeated by almost all sets in the corpus as shown in Fig. 6. In addition, the "own" set of HRTFs won more than 9 times against slightly more than half of the sets in the corpus with the exception of Listener 4 (Fig. 7).

3.3. Estimation of probability of winning a tournament

The probability of winning a tournament can be estimated from each probability of victories in each match obtained from round-robin comparison if 32 sets of HRTFs in the tournament are selected. Though a Swiss-style tournament is performed in DOMISO, the use of a normal-style tournament is assumed here. In such case, a specific set of HRTFs can win a tournament when it defeats five opponents. Here, P_{AB} is defined as the probability of set A defeating set B. The probability of winning a tournament, P_{Awin}, can be estimated as

$$P_{Awin} = P_{AB_1} P_{AB_2} P_{AB_3} P_{AB_4} P_{AB_5}, \qquad (2)$$

where a B_n set indicates an n-th opponent. All the combinations of five opponents from 130 candidates were calculated and P_{Awin} was averaged. The averaged probability of winning a tournament is shown in Table 5. Here, when assuming that an each evaluation was performed by 50%, P_{Awin} is 3% ($= 1/2^5$), while in Table 5, the mark "*" indicates that it is larger than 3%.

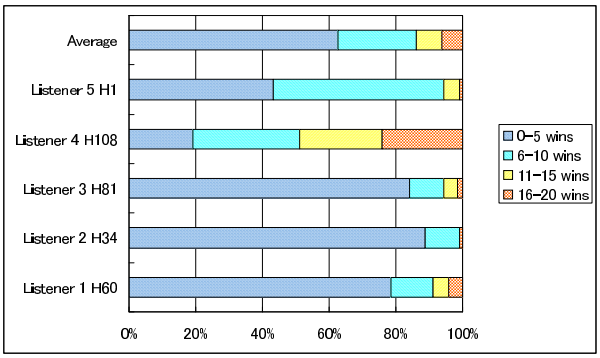


Figure 6: *Classification of all evaluation results in round-robin conducted twenty times ("worst" HRTFs).*

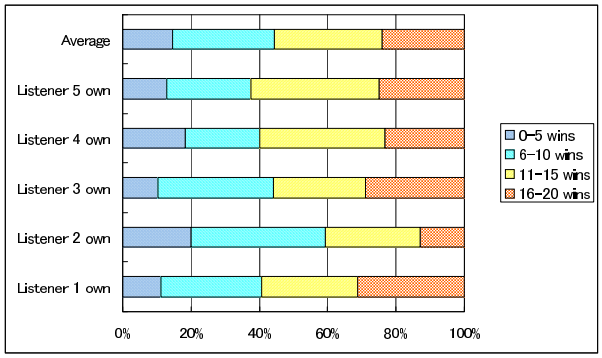


Figure 7: *Classification of all evaluation results in round-robin conducted twenty times ("own" HRTFs).*

From Table 5, almost all sets of HRTFs included as "best" HRTFs can be seen to have had a probability which greatly exceeded 3%, and the averaged probability of all five listeners was about 15%, which is five times 3%.

On the other hand, that of the "worst" HRTFs was below 3%, and the averaged probability was about 0.3 times 3%. Moreover, regarding "own" HRTFs, the probability of winning the tournament was about 6.4%.

Table 5: *Calculated percentage [%] of selected sets of HRTFs needed to win the tournament of each listener*

Listener	HRTFs				
	best1	best2	best3	worst	own
1	29.3*	42.6*	33.4*	0.01	8.6*
2	9.7*	19.2*	11.7*	0.04	2.4
3	21.7*	4.4*	0.7	0.02	8.5*
4	3.6*	10.0*	5.7*	4.0*	6.0*
5	4.3*	5.3*	5.8*	0.2	6.6*
Average	14.49*			0.85	6.42*

4. CONCLUSIONS

To examine the process of subjective selection of HRTFs in detail and to clarify the features of the individualization technique by subjective evaluation, an experiment based on DOMISO to choose the best set of HRTFs was repeated ten times. As a result, it was found that not only the individualization results but also each subjective evaluation were not definitive. A round-robin comparison of the 130 HRTFs in the corpus was also conducted. As a result, for the set of HRTFs which easily won in DOMISO, the average probability to win a tournament among 32 HRTFs was estimated to be about 15%. This probability greatly exceeds the probability (3%) when one set of HRTFs was selected from 32 candidates at random. Our results show that appropriate HRTFs can be chosen by subjective evaluation with high probability. In practical applications, it is important to control the perceived position of a sound image not only in the horizontal plane but also at arbitrary positions around a listener. We have previously found that the horizontal upper orbit is effective to control a particular sagittal plane [11]. Improvement of our individualization method will be attempted in our future work so that individualized HRTFs can perform well for all 4π directions with high stability.

5. ACKNOWLEDGEMENTS

This study was partially supported by a Grant-in-Aid for Young Scientists (B) from the Ministry of Education, Culture, Sports, Science and Technology of the Japanese Government (17700090). The authors wish to thank Ms. Y. Yang for her efforts in the experiments.

6. REFERENCES

[1] J. Blauert, "Spatial Hearing," The MIT Press, 1983.

[2] M. Morimoto and Y. Ando, "On the simulation of sound localization," *J. Acoust. Soc. Jpn. (E)*, vol. 1, pp. 167–174, 1980.

[3] E.M. Wenzel, M. Arruda, D.J. Kistler and F.L. Wightman, "Localization using non-individualized head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 94, pp. 111–123, 1993.

[4] Y. Suzuki, K. Watanabe, Y. Iwaya, J. Gyoba and S. Takane, "Adjustment of interaural time difference in head related transfer functions based on listeners' anthropometry and its effect on sound localization," in *Proc. of 149th Meeting of the Acoustical Society of America*, 3aPP10, p.2485, 2005.

[5] J.C. Middlebrooks, "Individual differences in external-ear transfer functions reduced by scaling in frequency," *J. Acoust. Soc. Am.*, vol. 106, pp. 1480–1492, 1999.

[6] K. Watanabe, K. Ozawa, Y. Iwaya, Y. Suzuki and K. Aso, "Estimation of interaural level difference based on anthropometry and its effect on sound localization," *J. Acoust. Soc. Am.*, vol. 122, pp.2832–2841, 2007.

[7] S. Sekimoto, R. Ogasawara, Y. Iwaya, Y. Suzuki and S. Takane, "Numerical investigation of effects of head sizes and ear positions on head-related transfer functions," in *Proc. of The Japan-China Joint Conference of Acoustics 2007*, (6 pages on CD-ROM), 2007.

[8] B.U. Seeber and H. Fastel, "Subjective selection of non-individual head-related transfer functions," in *Proc. of the 2003 International Conference on Auditory Display*, pp. 259–262, 2003.

[9] Y. Iwaya, "Individualization of head-related transfer functions with tournament-style listening test: Listening with other's ears," *Acoust. Sci & Tech.*, Vol. 27, No. 6, 2006.

[10] K. Watanabe, Y. Iwaya, J. Gyoba and Y. Suzuki, "Head-related transfer function corpus for virtual auditory environment," in *Proc. of the Forum on Information Technology 2005 (FIT2005), Information Technology Letters*, pp. 237–240, 2005 [in Japanese].

[11] Y. Iwaya, S. Yairi and Y. Suzuki, "Individualization of head-related transfer functions in virtual auditory display with listening-test," in *Proc. of The Japan-China Joint Conference of Acoustics 2007*, (6 pages on CD-ROM), 2007.